

Mathematik II für Biologen

Übungsblatt 10 (Abgabe am 20.06.2014)

Aufgabe 37

(10 Punkte)

Die Zufallsvariable X habe folgende Dichte:

$$f(t) = \begin{cases} c(1 - t^2) & \text{falls } 0 \leq t \leq 1 \\ 0 & \text{sonst} \end{cases}$$

- Skizzieren Sie $f(t)$ und bestimmen Sie die Konstante c so, dass f eine Dichte ist. (Hinweis: Die Fläche unter der Kurve muss 1 sein.)
- Berechnen Sie den Erwartungswert und die Varianz von X .
- Berechnen und skizzieren Sie die zugehörige (kumulative) Verteilungsfunktion $F(t)$.
- Bestimmen Sie den Median von X , d.h. die Zahl med mit $F(\text{med}) = \frac{1}{2}$.
- Bestimmen Sie die Wahrscheinlichkeit dafür, dass X einen Wert zwischen 0,4 und 0,8 annimmt, also $P[0,4 \leq X \leq 0,8]$. Zeichnen Sie diese Wahrscheinlichkeit in die Skizzen aus den Teilen a und c ein.
- Berechnen Sie die bedingte Wahrscheinlichkeit, dass $X \leq 0,2$, unter der Bedingung $X \leq 0,5$.

Aufgabe 38 MATLAB

(10 Punkte)

Die Datei `fishy.dat` enthält in der ersten Spalte die Länge (in cm), in der zweiten Spalte das Gewicht (Einheit leider unbekannt) und in der dritten Spalte den DDT-Gehalt (in ppm) von $n = 96$ Welsen, die im Tennessee River in Alabama, USA, gefangen wurden (Quelle: Mendenhall, Sincich: *Statistics for Engineering and the Sciences*, Prentice Hall. Appendix III. Daten leicht geändert, daher das y im Dateinamen.).

- Bestimmen Sie den empirischen Median der DDT-Gehalte.
- Ist dieser empirische Median signifikant (zum Signifikanz-Niveau $\alpha = 5\%$) von 10 verschieden? Beantworten Sie diese Frage mit einem zweiseitigen Vorzeichentest.
- Bestimmen Sie das zum Vorzeichentest gehörige 95%-Vertrauensintervall $[a, b]$ für den "wahren, theoretischen" Median des DDT-Gehaltes eines solchen Fisches, indem Sie den Test aus Aufgabe (b) für verschiedene Werte wiederholen und dabei beobachten, ob der Test verwirft. Bestimmen Sie dabei a und b so genau, dass die erste Ziffer nach dem Komma sicher stimmt.

Unvollständiger MATLAB-Code:

```
load fishy.dat
ddt=fishy(:,3);
[p,h]=signtest(ddt,10) % help signtest
```

Aufgabe 39

(3 Zusatzpunkte)

Üben Sie bis spätestens 29.06.14 auf www.khanacademy.org die *Skill*

- *Empirical rule.*

HINWEISE: Um die Aufgaben zu lösen, muss man ausschließlich mit den folgenden Werten für die Normalverteilung arbeiten: Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, dann gilt

$$P[|X - \mu| \leq n\sigma] \approx 68\%, 95\%, 99,7\% \text{ für } n = 1, 2, 3.$$

Für allgemeine Hinweise zu Khan-Aufgaben siehe Aufgabe 3 (Blatt 1).

Aufgabe 40

(10 Punkte)

Die Inkubationszeit X (in Monaten) einer bestimmten ansteckenden Krankheit wird als *log-normal* verteilt mit $\mu = 0$ und $\sigma = 0,4$ modelliert, d.h. man nimmt an, dass $\log X \sim \mathcal{N}(0; 0,4^2)$, wobei \log der natürliche Logarithmus ist.

- Geben Sie die Wahrscheinlichkeit an, dass die Inkubationszeit mehr als 3 Monate beträgt.
HINWEIS: Der MATLAB-Befehl `normcdf(x)` berechnet $\Phi(x)$.
- Berechnen Sie den Median med von X .
- Für eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable Y gilt bekanntlich

$$P[\mu - \sigma \leq Y \leq \mu + \sigma] \approx 68\% \quad \text{und} \\ P[\mu - 1,96\sigma \leq Y \leq \mu + 1,96\sigma] \approx 95\%.$$

Finden Sie Konstanten c_1 und c_2 , so dass

$$P\left[\frac{\text{med}}{c_1} \leq X \leq \text{med} \cdot c_1\right] \approx 68\% \quad \text{und} \\ P\left[\frac{\text{med}}{c_2} \leq X \leq \text{med} \cdot c_2\right] \approx 95\%.$$

Aufgabe 41

(10 Punkte)

Die Forschergruppe *Publischi* betreibt Data-Mining:⁹ An einem großen Datensatz testen die Wissenschaftler 10 unterschiedliche Nullhypothesen auf dem Signifikanzniveau $\alpha = 5\%$. Eine dieser Nullhypothesen wird verworfen. Die Gruppe *Publischi* veröffentlicht einen Artikel über die zugehörige Alternativhypothese, die durch ihre Arbeit statistisch bewiesen wurde.

Wir nehmen an, dass die Nullhypothesen so unterschiedlich waren, dass die Ausgänge der 10 Hypothesentests unabhängig voneinander waren. Weiter nehmen wir an, dass die Wahrscheinlichkeit, eine wahre Nullhypothese zu verwerfen, jeweils 5% betrug. (Warum?)

- Falls alle Nullhypothesen wahr sind, wie groß war die Wahrscheinlichkeit, dass dennoch mindestens eine von ihnen verworfen wird?
- Kommentieren Sie kritisch das Vorgehen von *Publischi*.



xkcd.com/882/

⁹vgl. z.B. <http://de.wikipedia.org/wiki/Data-Mining>